

From Data to Business Discovery

Deriving Better Insights through the Exploration of 'Big Data'

Issue 1

2

From the Gartner Files:
From Data to Decision: Delivering
Value From 'Big Data'

6

QlikView and Big Data

7

QlikView's In-Memory Technology
and Big Data

8

QlikView Google BigQuery
Integration

10

QlikView and Hadoop: King.com
Case Study

11

Conclusion: QlikView Gives Big
Data Big Meaning

Featuring research from

Gartner



From Data to Decision: Delivering Value From ‘Big Data’

“Big data” initiatives create significant opportunities for business and IT leaders, but both must come to terms with the challenges they introduce. A focus on the impact of analytical outcomes in supporting material and measurable business decisions is as important as addressing technology issues.

Impacts

- Business demand for data to support fact-based decisions will focus IT attention on information management, analytics and technology.
- Failure to balance data volume and quality to support material and measurable decisions by both business and IT leaders will undermine the success of big data initiatives.
- Business leaders will feel threatened by analytical outputs, and find reasons to reject them, unless they develop high levels of trust in their pedigree and provenance.

Recommendations

- CIOs and IT leaders should seek to understand and exploit new technologies and analytical methods to enable extreme information management.
- CIOs and IT leaders must appreciate the relevance and business impact of analytical outcomes in supporting the key material and measurable business decisions.

- To know where to focus, CIOs and IT leaders must understand the decision being sought and the process in order to balance the need for data volume against quality.
- Focus on delivering relevant input to the decision-making process. Sometimes volume represents value; other times, it is just noise.
- CIOs and IT leaders should pay careful attention to ensure that business leaders understand and trust the provenance and pedigree of analytical outcomes or risk having them disregarded.
- Both IT and business leadership should be aware of the dangers of human bias and behavior to acceptance of analytical outcomes in management decision making.

Strategic Planning Assumption(s)

Through 2015, more than 85% of Fortune 500 organizations will fail to exploit big data for competitive advantage.

Analysis

The confluence of current trends in social computing, mobility and cloud computing, combined with new sources and types of information, is creating a powerful nexus of forces that will impact every organization. As a result, CIOs and IT leaders are facing the reality of addressing the challenges of extreme information management (popularly referred

to as “big data”) to deliver against the increasing business demands for better data with which to make better decisions. However, merely collecting and analyzing data (including taking advantage of data from open sources) is not enough to ensure success. The analytical output must be presented in a timely and appropriate fashion so that decisions made as a direct consequence have a material and measurable impact on the productivity, profitability or efficiency of the organization. While the technical issues surrounding the timely collection and analysis of data present significant technical challenges, business leaders must learn to trust the provenance and pedigree of the analytical output, evolve existing management decision-making processes and overcome any personal bias such that decisions made take full advantage of the resources used to derive the analytical inputs.

As we progress from historical analysis (hindsight), through understanding (insight), toward predictive analytics (foresight), the ultimate goal is to create an unbroken chain from data through decision, where there is a clear and unequivocal relationship between the two (line of sight). Any failure at any point along this chain in either technology- or business-related processes represents a failure to deliver the benefits, and risks big data initiatives being branded a failure. At this point in time, we believe that most organizations are ill-

prepared to address both the technical and management challenges posed by extreme information management, and as a direct result, only a few (Type A) CIOs will be able to effectively exploit this significant trend for competitive advantage. Others should temper hype-driven business enthusiasm with a realistic assessment of existing skills and capabilities to maximize their prospects for success, and may be advised to get the basics (staffing, skills and existing information management architectures) fully up to speed before taking the plunge.

Impact: Business demand for data to support fact-based decisions will focus IT attention on information management, analytics and technology

The long-standing desire by business leaders for better data with which to make better, fact-based decisions places great pressure on CIOs and IT leaders to deliver. The current nexus of forces (social, mobile, cloud and information) has radically increased the volume, variety and velocity with which data is being presented, leading to significant technical challenges

in collecting, storing, processing, analysis and visualization of data. The imminent and ongoing development of new technologies to support faster processing (such as in-memory computing), new processing architectures (such as Hadoop and SAP's HANA) and new analytical tools (such as IBM's Watson and predictive analytics in general) provide a rich and potentially inexhaustible toolbox of technology options with which to attack the challenges of extreme information management. The problem is that, faced with the

Figure 1 | Impacts and Top Recommendations for Delivering Value From Big Data

| Impacts | Top Recommendations |
|---|--|
| Business demand for data to support fact-based decisions will focus IT attention on information management, analytics and technology. | <ul style="list-style-type: none"> CIOs and IT leaders should seek to understand and exploit new technologies and analytical methods to enable extreme information management. CIOs and IT leaders must appreciate the relevance and business impact of analytical outcomes in supporting the key material and measurable business decisions. |
| Failure to balance data volume and quality to support material and measurable decisions by both business and IT leaders will undermine the success of "big data" initiatives. | <ul style="list-style-type: none"> To know where to focus, CIOs and IT leaders must understand the decision being sought and the process in order to balance the need for data volume against quality. Focus on delivering relevant input to the decision-making process. Sometimes volume represents value; other times, it is just noise. |
| Business leaders will feel threatened by analytical outputs and find reasons to reject them, unless they develop high levels of trust in their pedigree and provenance. | <ul style="list-style-type: none"> CIOs and IT leaders should pay careful attention to ensure that business leaders understand and trust the provenance and pedigree of analytical outcomes or risk having them disregarded. Both IT and business leadership should be aware of the dangers of human bias and behavior to acceptance of analytical outcomes in management decision making. |

Source: Gartner (March 2012)

substantial but relatively familiar challenges posed by the adoption and deployment of new information technologies and architectures, CIOs, IT leaders, and enterprise and information architects will focus almost exclusively on these familiar, albeit challenging, areas and avoid addressing the much less familiar business issues that are equally critical to overall success.

Recommendations

CIOs and IT leaders should seek to understand and exploit new technologies and analytical methods to enable extreme information management.

As the recognized IT authority inside the organization, CIOs must explore, evaluate and, where appropriate, deploy relevant new technologies to meet their mission to deliver valuable analytical outputs from the growing volume and variety of data sources available to the organization. This will involve significant expense, both in terms of technology and the deployment of resources, and will thus require existing resources to be diverted or additional resources made available. Consider the creation of a competency center within the organization to concentrate resources and skills. Success will require an understanding of the competitive advantage and business benefits to be gained, which necessitates a more active engagement in, and understanding of, the nontechnical elements of the extended management decision-making

process. Gartner's Business Analytics Framework is a helpful tool to build this knowledge and foster the right discussions with business leaders.

CIOs and IT leaders must appreciate the relevance and business impact of analytical outcomes in supporting the key material and measurable business decisions.

CIOs and IT leaders must divide their attention between the familiar challenges of technology assessment and deployment, and the less familiar challenges of the business processes (formalized or implicit) involved in management decision making. Critical to this will be their ability to understand and focus on the business impact of analytical outcomes, and to prioritize (most probably in close cooperation with their business colleagues) and identify the most valuable, significant or critical processes where the return on analytical investment will bring the greatest or most immediate returns. The focus on material and measurable decisions is critical – the decisions must mean something significant to the business, and the net benefits of the investments must be measurable in such a manner that a clear relationship between the data, subsequent analysis, decision and outcome is established and evident. Focusing on analysis to improve and enhance decisions around the most valuable processes (as defined by the business leadership) is a good place to start. Failure to focus on the most important business areas will be the most common cause of failure in big data initiatives.

Impact: Failure to balance data volume and quality to support material and measurable decisions by both business and IT leaders will undermine the success of big data initiatives

There is a challenging paradox in the balance between volume and quality of data, and their value in delivering insightful input into the decision-making process. With modern analytical engines and architectures, there are times when increasing the volume of data processed improves the value of the outcome. A good example is IBM's Watson. Its language-processing algorithms look for patterns, so the more data it has to analyze, the more accurate, refined and sensitive those algorithms become. Simply put, Watson gets better the more data it has.

At the opposite extreme however, concerns over accuracy and reliability of data matter more than the volume. Indeed, the accuracy of the analytical outcome may be significantly adversely impacted if any volume of input is inaccurate (reflecting quality) or even fraudulent. In these circumstances, more data (especially if it is of questionable or unknown quality) simply becomes noise to mask the actual result being sought from the smaller volume of accurate, high-quality data. Hence, it is important to understand the probable quality of the dataset being processed and the type of decision (a trend based on overall sentiments, perhaps, versus an absolute based on quantitative records) being sought.

Recommendations

To know where to focus, CIOs and IT leaders must understand the decision being sought and the process in order to balance the need for data volume against quality.

The paradoxical relationship between volume and quality leaves information architects and IT leaders with a challenge. They must determine, on a case-by-case basis almost, those instances where the value of greater volume of data outweighs the cost of collecting it and processing it, and to what extent that volume outweighs inherent quality issues. This requires a deeper understanding of the decision-making process and the type of decision being made.

Focus on delivering relevant input to the decision-making process. Sometimes volume represents value; other times, it is just noise.

More data does not automatically create greater accuracy and, in some cases, actually introduces such a level of noise that the weaker patterns being sought are obscured. This principal is well-established in existing research on human factors in information overload situations (such as experienced in a battlefield/combat situation or on a busy financial trading floor). The emphasis must be on providing the information required to take the relevant decision and action, and avoid overloading the decision maker with unnecessary or distracting information. There is also strong support for the notion of “decision paralysis,” where the knowledge that additional information is (or will shortly be)

available leads the less-experienced human subject to postpone a decision until it arrives, thus entering a perpetual loop in which the decision is infinitely postponed.

Impact: Business leaders will feel threatened by analytical outputs, and find reasons to reject them, unless they develop high levels of trust in their pedigree and provenance

In an increasingly digital world, it is important to differentiate between fully automated (usually operational/tactical) decisions where the decision response is automatically triggered by the analytical input and the (generally more complex) strategic decisions where the analytical input is more likely to be a decision support tool to the human responsible for taking the actual decision. A good example of the former would be UPS’s Package Flow system, which allows UPS to automate the planning of daily delivery routes. The system obviates the need for human expertise and knowledge of the exact whereabouts of a destination address, and delivers not only greater accuracy, but also significant improvements in efficiency. Such routine tactical processes benefit from the fully automated approach, but care must be taken to ensure that they are only applied to closed, well-understood systems that are clearly defined and rigidly delineated. Similar automated approaches to areas such as financial trading (where the economic environment is a vastly more complex system) are blamed by many for creating increased instability and volatility in markets.

Secondly, due consideration must be given to the human element in decision making. Most organizations still reflect a very hierarchical structure in which (major) strategic decision making is dominated by a small number of senior business leaders. These individuals will generally exhibit (to a greater or lesser extent) a series of behavioral biases, personal prejudices or simply varying attitudes to risk that will influence their decision making in indeterminate and unpredictable ways. Limited understanding of technology may make them mistrustful of an analytical “black box” approach, even though it may represent a more structured, balanced and accurate assessment of risks, and produce a better decision than their more subjective view based on experience and other (indeterminate) factors.

Recommendation

CIOs and IT leaders should pay careful attention to ensure that business leaders understand and trust the provenance and pedigree of analytical outcomes, or risk having them disregarded.

Trust and understanding are the critical elements here. A lack of understanding of where the answer came from and the accuracy, pedigree and provenance of the underlying data can lead to a lack of trust in the quality of the analytical outcome, which, in turn, can lead to rejection and the view that IT failed to deliver. In some cases, such mistrust may be deep-seated and difficult to overcome, and will require significant reassurance

(and perhaps parallel testing and evaluation over a period of time, comparing the analytical prediction against the actual outcome and the conventionally derived decision), and may well entail significant delay to the implementation project. A variety of reasons for the failure to adopt the analytical approach (some of which may be reasonable, and others less so) are likely to be encountered and will require sensitive and objective responses to address them one by one.

Both IT and business leadership should be aware of the dangers of human bias and behavior to acceptance of analytical outcomes in management decision making.

Human behavior and personal bias, especially in senior business decision makers, can be a powerful driver to undermine the introduction and success of big data initiatives. Lack of trust, limited understanding or fear that it undermines their position, authority and reputation inside the organization are powerful drivers for an individual to reject the analytical inputs to the decision-making process. Under such circumstances, the realignment of analytical systems as a “decision support” mechanism to the management decision-making process, provides valuable, accurate and timely information, while leaving the actual decision in the hands of the individual, is a circumspect approach.

Over time, subject to continuing accuracy, the value of the solution will become readily apparent, and only in exceptional circumstances (often as a result of situations that were not visible or accessible to the data collection process) will a contrarian decision be made.

Nevertheless, the sensitivity of these issues, combined with the significant impact they can have on the perceived success of the initiative, means that they must be adequately and seriously addressed by the CIO and IT leadership throughout the planning, execution and operational phases of any significant big data project.

Source: Gartner RAS Core Research, G00232634, P. Dawson, 28 March 2012

QlikView and Big Data

Business users are constantly being challenged to efficiently access, filter, and analyze data – and gain insight from it – without using data analytics solutions that require specialized skills. They need better, easier ways to navigate through the massive amounts of data to find what’s relevant to them, and to get answers to their specific business questions so they can make better business decisions more quickly.

The growth in adoption of massively parallel processing solutions for handling ever larger volumes of data – whether structured or unstructured – is driving demand for analysis tools to enable business users to derive insights from Big Data.

QlikView takes a two-pronged approach to this challenge.

First, QlikView’s approach has always been to understand what it is that business users require from their analysis, rather than to force-feed a solution that might not be appropriate. Providing appropriate data for the use case is more valuable to users than providing all the data, all the time. For example, local bank branch managers may want to understand the sales, customer intelligence, and market dynamics in their branch catchment area, rather than for the entire nationwide branch network. With a simple consideration like this, the conversation moves from one of large data to one of relevance and value.

In any organization, the number of people who need to analyze extremely large data volumes is typically relatively small. For example, a retail bank might have thousands of branches; however, there may be only a hundred business analysts in a centralized, corporate role. While branch managers only need slices of data that are relevant to their operations, the corporate analysts may need access to much larger data volumes.

QlikView is designed to accommodate both types of use cases and enables users to focus on the data that is relevant to them and is of the highest value to them and their area of interest. By taking appropriate slices of the data – big or small – QlikView serves as an analytical app platform downstream of the data sources, to

provide business analysts and less technical business users alike the insight they need from the data that is most relevant to them.

Second, QlikView has been addressing, and continues to address, the Big Data challenge by ensuring that targeted QlikView apps can address the amounts of data that are needed to ensure the relevancy of the app for business users. Here's how:

- **Open Data Protocol:** QlikView provides an open data protocol (QVX, or QlikView data exchange) via a series of APIs (application programming interfaces) developers use to interface with the APIs of Hadoop- and other Big Data system providers. QlikView's QVX protocol can be used to connect to Hadoop based systems via two different methods:
 - Disk-based QVX file extracts from Hadoop (push)
 - "Named pipe" QVX connector for Hadoop (pull)

- **QlikView Google BigQuery Integration:** QlikView has partnered with Google to provide a visual analytics front end to the Google BigQuery solution. The QlikView Google BigQuery integration solution enables non-technical and non SQLsavvy users to interact effectively with billions of rows of data in seconds to find what is relevant to them and ask their own questions on the BigQuery data.
- **In-Memory Technology:** Recent trends in large memory available on standard Intel hardware allow QlikView to handle ever-larger volumes of data in memory (which provides users with a super-fast, interactive experience).
- **3rd Party Connectors:** QlikTech has established partnerships with third-party providers to connect with Big Data sources such as Attivio, DataRoket, and Informatica. A

QVX SDK is available to all third-party developers who wish to build custom connectors for any system with an open API.

- **Best Practices:** QlikView best practices promote an architecture-led deployment when handling very large data volumes, such as making proper use of distributed servers in a clustered environment; constructing appropriate apps for the intended audience; using sophisticated data reload engines; and using document chaining where necessary to allow aggregated views to be coupled with detail-level views while optimizing hardware resources.
- **Strategic Investments:** In June 2012 QlikTech acquired Expressor Software and now offers the QlikView Expressor Server, which provides a metadata intelligence capability and advanced data integration capabilities.

Source: QlikView

QlikView's In-Memory Technology and Big Data

QlikView is known as a pioneer in in-memory BI. In-memory technology is important for working with Big Data sources, because it enables on-the-fly, deep-dive analytics on this ocean of data, without business users needing to be concerned about or aware of the complex, back-end data integration taking place in the background.

QlikView's in-memory technology makes the Big Data experience more powerful and valuable because it:

- **Holds data in memory for multiple users, for a super-fast user experience.** QlikView holds all the data needed for analysis in memory, where it is available for immediate exploration by users. Users experience zero wait time as QlikView performs the calculations needed to deliver the aggregations users request. The "rocket science" is in getting the calculations done quickly. QlikView is a multi-user, distributed environment; it stores common calculations and

shares them among users, so the calculations don't have to be redone every time someone needs them.

- **Maintains associations in the data automatically.** QlikView's inference engine enables the green / white / gray associative experience. This engine automatically maintains the associations among every piece of data in the entire data set used in an application – neither developers nor end users have to maintain the associations. As a result, users

aren't limited to static reports, pre-determined drill paths, or pre-configured dashboards. Instead, they can navigate their data up, down, and sideways, exploring it any way they want to.

- **Calculates aggregations on the fly as needed.** QlikView's inference engine calculates aggregations on the fly based on selections the user makes (which we call the "state" of the app). As a result, users aren't

limited to predefined calculations (and hence preconceived insights based on data joins made by IT). Users can define whatever view or type of insight they want and QlikView dynamically calculates the answer. QlikView only calculates the aggregations the user asks for; it does not pre-calculate aggregations the old queries-and-cubes way. It processes calculations instantaneously, as they are needed.

- **Optimizes the power of the processor.** QlikView distributes calculations across all available cores to maximize the performance experienced by the user. Unlike technologies that simply "support" multi-processor hardware, QlikView is optimized to take full advantage of all the power of multi-processor hardware, thereby maximizing performance and the hardware investment.

Source: QlikView

QlikView Google BigQuery Integration

Google BigQuery is a web service that brings big data analytics to all businesses via the cloud. With Google BigQuery, developers and business users can quickly and easily gain business insights from massive amounts of data without any hardware or software investments. Scalable and easy to use, BigQuery lets developers and businesses tap into powerful data analytics on demand.

QlikView and Google BigQuery integration allows users to access big data in real time and derive value and visualize the information. The goal of the integration is to provide seamless Business Discovery capabilities for the business users on very large data

sets by leveraging the value of the BigQuery platform. With QlikView's unique associative experience, business users can navigate through massive amounts of data to find what's relevant to them, and to get answers to their specific business questions without requiring technical skills.

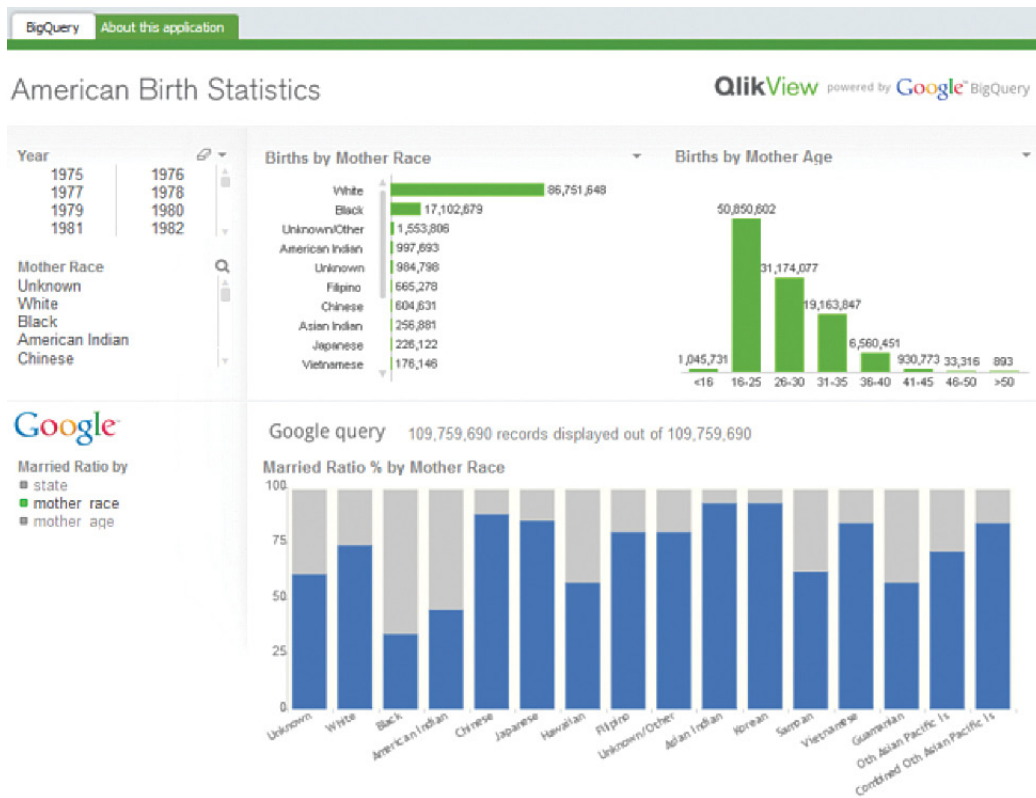
The integration leverages two QlikView capabilities;

- **A custom connector:** QlikView developers can use the QlikView BigQuery connector to load BigQuery data into QlikView's in-memory data model so business users can remix and reassemble it in new views and

create new visualizations on the fly. Users can make selections in the data and see what data is associated, and what data is not.

- **An extension object:** For massive sets that are too big to fit in memory, even when compressed, QlikView developers can create an extension object to directly query the BigQuery platform. Business users can interact with the BigQuery data by making selections in list boxes to get just the relevant cut of the data they need in a user-friendly chart or graph, without creating a single line of SQL code.

Figure 2 | QlikView Google BigQuery demo application (<http://qlikview.com/bigquery>)



Source: QlikView

The technical details of the integration are provided in the [QlikView Google BigQuery Integration white paper](#). A QlikView demo app, showing the integration, is available on the QlikView demo web site (<http://qlikview.com/bigquery>). An example extension object and the QlikView BigQuery custom connector are available to download for free on the QlikMarket web site (<http://market.qlikview.com/>).

Business users are constantly being challenged to efficiently access, filter, and analyze data, and gain insight from it without using data analytics solutions, which require specialized skills.

QlikView Google BigQuery integration, enables QlikView customers and partners to build Business Discovery solutions that take advantage of the computing power,

scalability, and security of Google's cloud platform. The integration allows navigating through the massive amounts of data to find what's relevant to the business users, and to get answers to their specific business questions so they can make better business decisions more quickly.

Source: QlikView

QlikView and Hadoop: King.com Case Study

King.com is an online gaming property based in Scandinavia. King.com is using QlikView with a Hadoop-based Big Data system to provide business users in the marketing function with rapid insight into customer behaviors captured through their use of the games. Everything from customer browsing activity while on the site to their interactions within each game played and many, many more metrics are captured in an on-premise Hadoop-based deployment. QlikView sits on top of the Hadoop

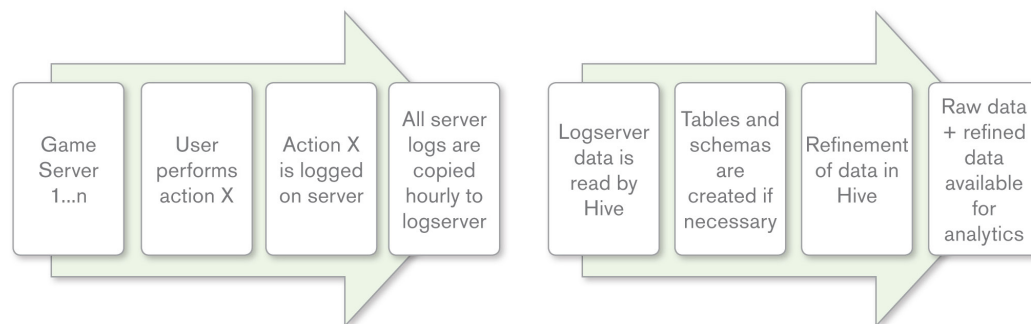
system to provide Business Discovery capabilities to enable King.com to more effectively target new customers, new games, new offers, and so on.

The data volumes acquired are impressive: 1.6 billion new rows are produced each day and stored in the Hadoop system. King.com uses QlikView to provide analysis of aggregated data from the Hadoop system. In their words, having billions of permutations of data reduces the statistical impact of the aggregation effect so that when they are analyzing

211 million rows in QlikView, they have a high degree of confidence that the data is highly representative of the entire data set within Hadoop.

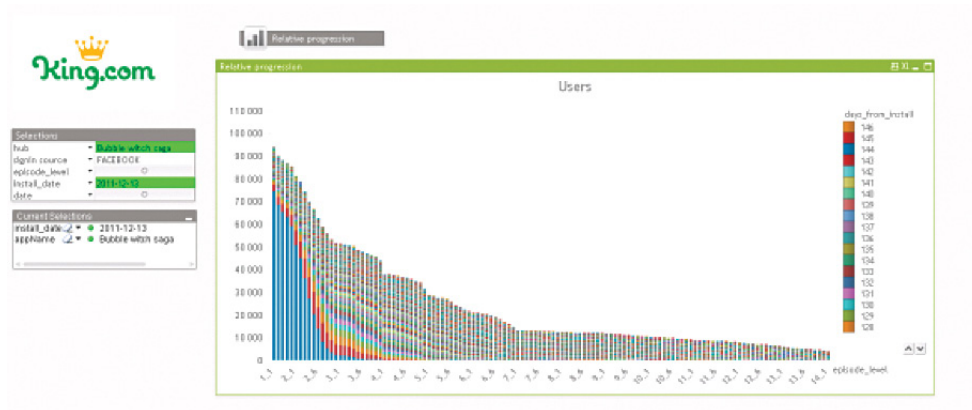
King.com utilizes a single 8 machine cluster to host their Hadoop environment. Each user event is logged, processed, and ultimately made available to QlikView users for analysis via an ODBC connector to Hive. The flow of data from the source gaming systems all the way to the QlikView analytics system is shown in Figure 3.

Figure 3 | *Flow of data in the King.com implementation of QlikView and Big Data*



Source: QlikView

Figure 4 | Screenshot of the King.com QlikView application



Source: QlikView

Conclusion: QlikView Gives Big Data Big Meaning

Making Big Data work for businesses involves a lot more than processing, unifying and delivering that data to the enterprise. Giving this deep ocean of data deep meaning is the key to making it work for the business. That's where QlikView delivers significant value. The

QlikView Business Discovery platform provides the ultimate Big Data front end for the enterprise. QlikView's core mission of "Simplifying decisions for everyone, everywhere" is also at the heart of its Big Data approach. It's about harnessing the real potential of Big Data and invisibly unleashing its value for users across the business.

Ultimately, business users don't care if it's "Big Data," "medium data" or "teeny-weeny data." What they care about is "Is it relevant data?" and "Is it valuable data that helps me solve my complex business problems?" The QlikView Business Discovery platform offers a Big Data solution that delivers a resounding "yes" to both questions.

Source: QlikView

QlikTech (NASDAQ: QLIK) is a leader in Business Discovery – user-driven Business Intelligence (BI). Its QlikView Business Discovery solution bridges the gap between traditional BI solutions and inadequate spreadsheet applications. The in-memory associative search technology QlikTech pioneered created the self-service BI category, allowing users to explore information freely rather than being confined to a predefined path of questions. Appropriate from SMB to the largest global enterprise, QlikView’s self-service analysis can be deployed with data governance in days or weeks. The QlikView Business Discovery platform’s app-driven model works with existing

BI solutions, offering an immersive mobile and social, collaborative experience. Headquartered in Radnor, Pennsylvania, QlikTech has offices around the world serving approximately 26,000 customers in over 100 countries.

QlikView

QlikTech and QlikView are trademarks or registered trademarks of QlikTech or its subsidiaries in the U.S. and other countries. Other company names, product names and company logos mentioned herein are the trademarks, or registered trademarks of their respective owners.

To learn more, visit qlikview.com

From Data to Business Discovery is published by Qliktech. Editorial supplied by Qliktech is independent of Gartner analysis. All Gartner research is © 2012 by Gartner, Inc. All rights reserved. All Gartner materials are used with Gartner’s permission. The use or publication of Gartner research does not indicate Gartner’s endorsement of Qliktech’s products and/or strategies. Reproduction or distribution of this publication in any form without prior written permission is forbidden. The information contained herein has been obtained from sources believed to be reliable. Gartner disclaims all warranties as to the accuracy, completeness or adequacy of such information. Gartner shall have no liability for errors, omissions or inadequacies in the information contained herein or for interpretations thereof. The opinions expressed herein are subject to change without notice. Although Gartner research may include a discussion of related legal issues, Gartner does not provide legal advice or services and its research should not be construed or used as such. Gartner is a public company, and its shareholders may include firms and funds that have financial interests in entities covered in Gartner research. Gartner’s Board of Directors may include senior managers of these firms or funds. Gartner research is produced independently by its research organization without input or influence from these firms, funds or their managers. For further information on the independence and integrity of Gartner research, see “Guiding Principles on Independence and Objectivity” on its website, http://www.gartner.com/technology/about/ombudsman/omb_guide2.jsp.